Postal Regulatory Commission
Submitted 10/7/2015 3:38:57 PM
Filing ID: 93501
Accepted 10/7/2015

BEFORE THE
POSTAL REGULATORY COMMISSION
WASHINGTON, D.C. 20268-0001

PERIODIC REPORTING (PROPOSAL ELEVEN)	Docket No. RM2016-1

PETITION OF THE UNITED STATES POSTAL SERVICE REQUESTING INITIATION OF A PROCEEDING TO CONSIDER A PROPOSED CHANGE IN ANALYTICAL PRINCIPLES (PROPOSAL ELEVEN)
(October 7, 2015)

Pursuant to 39 C.F.R. § 3050.11, the Postal Service requests that the Commission initiate a proceeding to consider a proposal to change analytical principles relating to the Postal Service's periodic reports. The proposal, labeled Proposal Eleven, is discussed below, and in greater detail in the attachment.

Proposal Eleven seeks authorization to change the statistical estimator for revenue, pieces and weight for the digital letter mail sampling in the Origin-Destination Information System - Revenue, Pieces and Weight (ODIS-RPW) system first discussed in Docket RM2015-11. Beginning Q2 FY2016, the Postal Service proposes to replace ODIS-RPW system manual data collection at some letter Mail Exit Points (MEPs) with an automated selection of digital images selected from the incoming secondary Delivery Barcode Sequencing (DBCS) second pass operation. Mail processed on the second pass is termed Delivery Point Sequenced, or DPS mail. The digital images captured would provide the same information as manual data collection except for a few items as discussed in Docket No. RM2015-11. In this proposal, ODIS-RPW would replace the direct expansion estimator with a ratio estimator that utilizes national End-of-Run (EOR) machine counts. The digital letter mail estimates utilizing the ratio estimator applied to the digital letter mail sampling frame would be combined with direct expansion estimates from the non-digital sampling frame.

The materials presented with this proposal show mathematically how the proposed ratio estimator for the letter mail digital sampling frame outperforms the direct expansion estimator for First-Class Mail single piece volume and revenue. First-Class Mail single piece is the only significant category from ODIS-RPW that is used in RPW reporting.

The Postal Service requests that the Commission initiate a rulemaking proceeding pursuant to 39 C.F.R. § 3050.11 to consider this proposal.

Respectfully submitted,

UNITED STATES POSTAL SERVICE

By its attorneys:

Daniel J. Foucheaux, Jr.
Chief Counsel, Pricing and Product Support

____
Eric P. Koetting

475 L'Enfant Plaza, S.W.
Washington, D.C. 20260-1137
(202) 277-6333
October 7, 2015

Proposal Eleven

PROPOSED CHANGE IN THE ESTIMATION FORMULA FOR ODIS-RPW USED IN RPW REPORTING RELATING TO DIGITAL LETTER MAIL SAMPLING

OBJECTIVE:

This proposal seeks a change in the statistical point and variance estimation methodology for the ODIS-RPW system estimates used in the "Revenue, Pieces and Weight By Class and Special Services" (RPW) report relating to letter and card mailpieces that will be sampled digitally.

Beginning Q2 FY2016 (January 1, 2016), a portion of the letter Mail Exit Points (MEPs) for certain five-digit zones will be identified as eligible for sampling digitally (digital images). These images would be selected during the incoming second-pass delivery barcode sequence (DBCS) mail processing operation and transmitted to a central server for data recording. This process would replace manual data collection, which currently requires a data collector to travel to the test site and then randomly sample the mail from containers. Manual data collection by a data collector can be referred to as "live" ODIS-RPW testing. All characteristics of a live mailpiece can be collected from an image of that mailpiece except for weight and letter and card shape. [1]

This proposed estimation methodology will improve the product estimates used for RPW by reducing bias and significantly lowering the calculated coefficient of variation for the same sample size. The purpose of this proceeding is to conclusively demonstrate these advantages, rather than to quantify any change in the actual

[1] An alternative methodology for estimating shape and weight for digital testing was presented in Docket No. RM2015-11 and approved in Order No. 2739 (September 30, 2015).

estimates themselves. While some small change cannot be ruled out, there is at present no way to ascertain the direction in which any particular category is more likely to change. The only significant category affected is First-Class Mail single piece letters and cards.

BACKGROUND:

RPW Reporting

Regulatory reporting of revenue, pieces and weight is presented in the RPW report filed quarterly with the Postal Regulatory Commission (Commission), in accordance with Commission Rule 3050.25. The RPW system used to develop this report was discussed in detail in witness Pafford’s testimony (USPS-T-3) in Docket No. R2006-1. Revenue, pieces and weight data for Postal Service products are obtained through various source systems, one of which is the ODIS-RPW probability sampling system (Docket No. R2006-1, Library References USPS-LR-L-16 and USPS-LR-L-17).

ODIS-RPW Probability Sampling System

The ODIS-RPW system is a probability-based destinating mail sampling system used to support the Postal Service's many and varied business needs for mail revenue and volume information. ODIS-RPW primarily supplies official RPW estimates of revenue, volume and weight for single-piece stamped and metered indicia mail.

ODIS-RPW data collectors travel to randomly-selected Mail Exit Points (MEPs) on randomly-selected days, and randomly sample mail as it arrives at the delivery units. Container and mailpiece skip sampling procedures are applied to the mail containers. Data collectors record mail characteristics from sampled mailpieces, including revenue, pieces, weight, mail class, subclass, and indicia.

Current Sample Design

Currently, MEPs are grouped into ‘strata’ based on approximate average daily volumes for letters, flats and parcels as maintained in the MEP system for each MEP. The stratification process results in letter strata being formed as well as strata associated with flats and parcels. Within strata, MEP-days are selected for data collector sampling. A discussion of the stratification and sampling methodology can be found in Library Reference USPS-LR-L-14, Docket No. R2006-1.

Statistical Expansion

Each month and quarter, statistical estimates for products are created by strata and then summed across strata for national estimates sent to the RPW system. The following describes the ODIS-RPW statistical formula, which first starts out with the definition of the strata.

Current Stratification

Indices for the current estimator are defined below;

$j = Sample\ Area$

$h = Strata\ based\ on\ Reference\ Volumes\ for\ a\ SA\ j$

$k = Mail\ Exit\ Point\ (MEP) data\ for\ a\ given\ (j, h)$

$l = mailpiece\ for\ a\ give\ (j, h, k)$

Current Estimators

Let $\hat{t}$ be defined as the current production estimator for a given product

$$\hat{t} = \sum_{j=1}^{J} \sum_{h=1}^{H_j} \hat{t}_{jh}$$

Where $\hat{t}_{jh}$ is,

$$\hat{t}_{jh} = \frac{N_{jh}}{n_{jh}} \sum_{k=1}^{n_{jh}} \frac{\hat{x}_{jhk}}{n_{jhk}} \sum_{l=1}^{n_{jhk}} y_{jhkl}$$

and,

$$y_{jhkl} = \begin{cases} revenue, pieces\ or\ weight\ for\ mailpiece\ l\ for\ the\ product\ of\ interest, \\ 0, \quad otherwise \end{cases}$$

$\hat{t}$ is a direction expansion estimator, whose estimates are constructed from the inverse of the probabilities of selection. Then, the variance of the direct expansion estimator is,

$$V(\hat{t}) = \sum_{j=1}^{J} \sum_{h=1}^{H_j} \mathrm{V}(\hat{t}_{jh})$$

With,

$$V(\hat{t}_{jh}) = N_{jh}^2 \left(1 - \frac{n_{jh}}{N_{jh}}\right) \frac{S_1^2}{n_{jh}} + \frac{N_{jh}}{n_{jh}} \sum_{k=1}^{N_{jh}} \hat{x}_{jhk}^2 \left(1 - \frac{n_{jhk}}{\hat{x}_{jhk}}\right) \frac{S_2^2}{n_{jhk}}$$

For a fuller discussion of these formulae, please see the technical notes in Appendix A.

PROPOSAL:

In this proposal, we introduce a ratio estimator that utilizes machine End-of-Run (EOR) counts from the second pass DBCS operation, or from the same set of pieces from which the digital images are selected. In order to utilize the new estimator, the zones that have been defined for digital letter mail sampling are first stratified in a different manner than what is currently done. The next sections introduce the proposed stratification, and then move to the new estimators.

Proposed Stratification

For each MEP assigned to the digital letter mail sampling frame, auxiliary information was obtained as to what portion of its delivery points were business related. The auxiliary information is maintained by Delivery Operations by route and zone. Five strata were defined based on this proportion. The idea of the stratification is that the product mix would vary based on its business composition; effectively controlling the variation through stratification.

Indices for the proposed estimator are defined below;

$i = Business\ Delivery\ Points\ (BDP)$

$j = Sample\ Area$

$k = MEP\ day\ for\ a\ given\ (i, j)$

$l = mailpiece\ for\ a\ give\ (i, j, k)$

Proposed Estimators

Let $\hat{t}_{ratio}$ be defined as the proposed estimator for a given product

$$\hat{t}_{ratio} = \sum_{i=1}^{5} \hat{t}_{i,ratio}$$

where,

$$\hat{t}_{i,ratio} = \frac{\sum_{j=1}^{J} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} \frac{x_{ijk}}{n_{ijk}} \sum_{l=1}^{n_{ijk}} y_{ijkl}}{\sum_{j=1}^{J} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}} \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} x_{ijk}$$

and,

$n_{ijk} = number\ of\ pieces\ sampled\ in\ MEP - day\ k\ in\ strata\ (i, j),$

$x_{ijk} = EOR\ count\ for\ MEP - day\ k\ in\ strata\ (i, j),$

$$y_{ijkl} = \begin{cases} revenue, pieces\ or\ weight\ of\ mailpiece\ l\ for\ the\ product\ of\ interest, \\ 0, \quad otherwise \end{cases}$$

The value $\hat{t}_{ratio}$ can be termed the ratio estimator. Then, the variance of the ratio estimator is,

$$V(\hat{t}_{ratio}) = \sum_{i=1}^{5} V\left(\hat{t}_{i,ratio}\right)$$

where,

$$V\left(\hat{t}_{i,ratio}\right) = \left(\frac{X_i}{\hat{X}_i}\right)^2 \left\{\sum_{j=1}^{J} N_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right)\frac{S_k^2}{n_{ij}} + \sum_{j=1}^{J} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{N_{ij}} x_{ijk}^2 \left(1 - \frac{n_{ijk}}{x_{ijk}}\right)\frac{S_{kl}^2}{n_{ijk}}\right\}$$

For a fuller discussion of these formulae, please see the technical notes in Appendix A.

RATIONALE:

The improvement in the ratio estimator over the direct expansion estimator comes in the correlation between EOR counts and the value t being estimated, whether it be product revenue or pieces. A standard textbook by William Cochran[2] shows that the ratio estimator has improved variance characteristics when this correlation is greater than or equal to one-half the ratio of the coefficient of variations of the EOR counts to average product revenue or volume. Please see section 3.3 in the Appendix for a broader discussion of this rationale.

IMPACT:

The following table shows the analysis of the components of the above inequality; the correlation coefficient, and the coefficients of variation for the EOR and

[2] Cochran, William G. (1977), Sampling Techniques. 3rd ed., John Wiley & Sons (Hoboken, NJ), at page 157.

the estimates for First-Class single piece revenue and volume. The actual correlation is the left hand side of the above described inequality, and the 'critical' correlation is the right hand side of the above inequality.

Table: Comparisons of Critical and Actual Correlation by Period.

Period	Variable	Coefficient of Variation (CV)	Critical Correlation	Less than (<) or Greater than (>)	Actual Correlation
FY15 PQ1	End-of- Run (EOR)	0.0075			
	Volume	0.0095	0.3939	<	0.6823
	Revenue	0.0267	0.1406	<	0.1864
FY15 PQ2	End-of- Run (EOR)	0.0080			
	Volume	0.0104	0.3875	<	0.7059
	Revenue	0.0132	0.3047	<	0.4798
FY15 PQ3	End-of- Run (EOR)	0.0080			
	Volume	0.0105	0.3787	<	0.5202
	Revenue	0.0124	0.3221	<	0.4991

Note: The ratio estimator has a lower variance if the critical correlation is less than (<) the Actual correlation.

What we see from this table is that CVs for EOR tend to be smaller than the product CVs. Therefore, the ratio estimator outperforms the direct expansion estimator even if the correlation coefficient is less than 0.5. Translated into the relationship of the critical correlation and actual correlation, we see that in all cases the critical correlation is lower. That is, the ratio estimator lowers the variance of the estimate when compared with the direct expansion estimator.

While not shown, we expect a fifteen to twenty percent decrease in the CVs for First-Class letter and card volume, and approximately a five to ten percent decrease in the CVs for First-Class letter and card revenue.

For further detailed discussion of the comparative advantage of using the ratio estimator over the direct expansion estimator, please see section 3.3 of the Appendix.

The analysis is not exhaustive of all remaining single-piece letter and card categories coming from the digital sampling frame that are used in the RPW report. Additional categories of letter mail processed on the second-pass operation include Media/Library, US Postal Service Mail, Free Matter for the Blind, Priority Mail, and include Extra Services of Certified Mail, Return Receipts, and Other Domestic Ancillary Services. With two exceptions, the ODIS-RPW statistical estimates of these categories and extra services represent less than three-tenths of one percent of the RPW reported revenue, pieces and weight. The exceptions are US Postal Service mail and Free Mail, representing approximately twenty-five percent and five percent of the RPW report volume, respectively.

Appendix A: Technical Notes

1 Introduction

The purpose of this technical note is to underline differences between the current and the proposed methodologies for stratification and estimation, and to exhibit advantages of the latter methodology.

2 Stratification

2.1 Current Stratification

Currently, the postal mail population is partitioned geographically by sample areas (SA) and each SA is further subdivided into 15 to 20 strata according to reference volumes of letters, flats, and parcels (RefVol). Strata based on RefVol are nested within SA stratification. The following indices are used to describe the current production sampling methodology;

$$\begin{aligned} j &= \text{Sample Area (SA)}, \\ h &= \text{Strata based on Reference Volumes (RefVol) for a SA } j, \\ k &= \text{MEP day for a given } (j,h), \\ l &= \text{mailpiece for a given } (j,h,k), \end{aligned}$$

Index j is defined over a set $\{1,\ldots,J\}$ where J represents the number of SAs. RefVol strata indexed by h is defined over $h = 1,\ldots,H_j$ for a given $j \in \{1,\ldots,J\}$, and H_j represents the number of strata for SA j. Typically, H_j is a number between 15 and 20. Other Indices are defined for k over a set $\{1,\ldots,n_{jh},\ldots,N_{jh}\}$, and l over

$\{1, \ldots, n_{jhk}, \ldots, x_{jhk}\}$ where

$$
\begin{aligned}
N_{jh} &= \text{the number of MEP-days for stratum } (j,h), \\
n_{jh} &= \text{the number of tests allotted to stratum } (j,h), \\
x_{jhk} &= \text{volume (EOR count) for } k\text{th MEP day in stratum } (j,h), \\
n_{jhk} &= \text{the number of mailpiece sampled for } k\text{th MEP day in stratum } (j,h)
\end{aligned}
$$

By taking M_{jh} = the number of digital MEPs in stratum(j,h), we have $N_{jh} = M_{jh} \times$ (Delivery Days for a Period).

Without loss of generality, we arrange MEP days so that the first n_{jh} MEP days corresponds to sampled MEP days and the rest of $(N_{jh} - n_{jh})$ non-tested MEP days. Similarly for index l, the first n_{ijk} mailpieces are assumed sampled and the rest of the mailpieces in $\{n_{ijk} + 1, \ldots, x_{ijk}\}$ are non-sampled mailpieces.

Index h is only relevant to the current stratification as reference volumes are not used under the proposed stratification for digital MEP described in section 2.2. Independently for each SA, a cluster algorithm classifies MEPs in homogenous groups. Approximately 3,000 independent strata in total are created nationwide by the algorithm, and estimation is done independently for each stratum using the sample gathered for that stratum.

Because of the sheer number of strata, sample sizes may be small at the strata level. Since strata are defined uniquely for each SA, there is no coherent way to substitute a sample from one stratum for another that is afflicted with a small sample size. The issue is alleviated if strata were defined uniformly across all SAs; then samples from the same strata across SAs can be pooled in producing more stable estimates. This is the rationale behind the change described in 2.2.

2.2 Proposed Stratification

A delivery point is categorized as either residential or business, and the numbers of residential and business delivery points are available for each digital MEP. We utilize the information through stratification in an attempt to establish an efficient sampling and estimation system.

Proportions of business delivery points (BDP) to total delivery points are computed for all digital MEPs and divided into 5 groups based on BDP values. Also, digital MEPs are geographically partitioned by SA. Consequently, BDP and SA cross-classify digital MEPs into $5 \times J$ strata where J is the number of SAs. Indices for the proposed stratification are defined below;

$$\begin{aligned} i &= \text{BDP category}, \\ j &= \text{SA}, \\ k &= \text{MEP day for a given } (i, j), \\ l &= \text{mailpiece for a given } (i, j, k), \end{aligned}$$

Indices are defined over sets; $\{1, \ldots, 5\}$ for i, $\{1, \ldots, J\}$ for j, $\{1, \ldots, n_{ij}, \ldots, N_{ij}\}$ for k, and $\{1, \ldots, n_{ijk}, \ldots, x_{ijk}\}$ for l. Cross-classification by BDP and SA means that collapsing of strata over index i or j would not affect the other stratification. This provides us with flexibility in choosing the estimator most appropriate to our situations.

The symbols can be interpreted in the same manner as they are in section 2.1 except for difference in stratification indices. For completeness, n_{ij} is the number of tests allotted to (MEP-days tested for) stratum (i, j), N_{ij} is the number of MEP-days for stratum (i, j), n_{ijk} is the number of mailpiece for kth MEP day in stratum (i, j), and x_{ijk} is the EOR count for kth MEP day in stratum (i, j). By letting $M_{ij} =$ the number of digital MEPs in stratum(i, j), we have $N_{ij} = M_{ij} \times (\text{Delivery Days})$.

2.3 Comparative Advantage

First, the boundaries of BDP stratification are defined uniformly across all SAs and this allows us to aggregate samples over SAs in the estimation stage. Such aggregations are not possible in the current stratification mechanism.

Reference volumes of letters, flats, and parcels (RefVol) are useful when the objective is to classify MEPs into letters, flats, and parcel groups. Since digital MEPs replaces MEPs currently dealing with those letters, RefVol would not provide any useful information in stratifying digital MEPs. On the other hand, considering that mail mixture could differ markedly between residential and business deliveries, BDP makes it possible to form homogenous strata. For instance, digital MEPs with higher proportions of business delivery points could have higher proportions of first-class single pieces compared to Digital MEPs that mostly focus on residential deliveries. The proposed stratification with BDP, therefore, will lead to a more representative sample and precise estimates.

3 Estimation Methodology

For simplicity, we limit our discussion of changes in estimation methodologies to "first-class single piece (1-C SP)," but the same methodologies apply to the other products. More specifically, we first demonstrate how "total revenue of 1-C SP" is estimated in the current production system.

3.1 Current Estimator

Let t denote "total 1-C SP revenue" and $\hat{t}$ be the current production estimator of t, then we have

$$\hat{t} = \sum_{j=1}^{J} \sum_{h=1}^{H_j} \hat{t}_{jh} \tag{3.1}$$

where expansion estimator of 1-C SP revenue for stratum h of SA j is

$$\hat{t}_{jh} = \frac{N_{jh}}{n_{jh}} \sum_{k=1}^{n_{jh}} \frac{\hat{x}_{jhk}}{n_{hjk}} \sum_{l=1}^{n_{jhk}} y_{jhkl} \tag{3.2}$$

for $j = 1, \ldots, J$ and $h = 1, \ldots, H_j$. The notation used above is presented in section 2.1 except for y_{jhkl}, which is defined here as

$$y_{ijkl} = \begin{cases} \text{revenue of mailpiece } l, & \text{if } l \text{ in test } k \text{ is a 1-C SP,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that End-of-Run (EOR) piece counts for SA j, stratum h of kth MEP-day, x_{jhk} are not readily available under the current configuration of MEPs consisting of multiple delivery zip codes. We estimate the total piece count by $\hat{x}_{jhk} = n_{jhk} \times$ (Container Skip) $\times$ (Piece Skip) where n_{jhk} denotes the number of pieces sampled for test k. The use of volume estimates $\hat{x}_{jhk}$ rather the the actual volume x_{jhk} potentially adds bias as well as uncertainty to an otherwise unbiased estimator. This is addressed with the proposed estimator discussed in section 3.2.

An approximate variance of $\hat{t}$ is of the form

$$V\left(\hat{t}\right) = \sum_{j=1}^{J} \sum_{h=1}^{H_j} V\left(\hat{t}_{jh}\right) \tag{3.3}$$

with

$$V\left(\hat{t}_{jh}\right) = N_{jh}^2 \left(1 - \frac{n_{jh}}{N_{jh}}\right) \frac{S_1^2}{n_{jh}} + \frac{N_{jh}}{n_{jh}} \sum_{k=1}^{N_{jh}} \hat{x}_{jhk}^2 \left(1 - \frac{n_{jhk}}{\hat{x}_{jhk}}\right) \frac{S_2^2}{n_{jhk}} \tag{3.4}$$

for $j = 1, \ldots, J$ and $h = 1, \ldots, H_j$ and

$$S_1^2 = \frac{1}{N_{jh} - 1} \sum_{k=1}^{N_{jh}} \left(t_{jhk} - \bar{t}_{jh}\right)^2$$

$$\bar{t}_{jh} = \frac{1}{N_{jh}} \sum_{k=1}^{N_{jh}} t_{jhk}$$

$$S_2^2 = \frac{1}{\hat{x}_{jhk} - 1} \sum_{l=1}^{\hat{x}_{jhk}} \left(y_{jhkl} - \bar{y}_{jhk}\right)^2$$

$$\bar{y}_{jhk} = \frac{1}{\hat{x}_{jhk}} \sum_{l=1}^{\hat{x}_{jhk}} y_{jhkl}$$

By substituting statistics for unknown parameters above, we obtain the variance estimator

$$\hat{V}\left(\hat{t}\right) = \sum_{j=1}^{J} \sum_{h=1}^{H_j} \hat{V}\left(\hat{t}_{jh}\right),$$

with

$$\hat{V}\left(\hat{t}_{jh}\right) = N_{jh}^2 \left(1 - \frac{n_{jh}}{N_{jh}}\right) \frac{s_1^2}{n_{jh}} + \frac{N_{jh}}{n_{jh}} \sum_{k=1}^{n_{ij}} \hat{x}_{jhk}^2 \left(1 - \frac{n_{jhk}}{\hat{x}_{jhk}}\right) \frac{s_2^2}{n_{jhk}}$$

where

$$s_1^2 = \frac{1}{n_{jh} - 1} \sum_{k=1}^{n_{jh}} \left(t_{jhk} - \hat{\bar{t}}_{jh}\right)^2$$

$$\hat{\bar{t}}_{jh} = \frac{1}{n_{jh}} \sum_{k=1}^{n_{jh}} \hat{t}_{jhk}$$

$$s_2^2 = \frac{1}{n_{jhk} - 1} \sum_{l=1}^{n_{ijk}} \left(y_{jhkl} - \hat{\bar{y}}_{jhk}\right)^2$$

$$\hat{\bar{y}}_{jhk} = \frac{1}{n_{ijk}} \sum_{l=1}^{n_{jhk}} y_{jhkl}$$

3.2 Proposed Estimator

For digital MEPs, EOR machine counts provide us with the reliable piece counts of letters and cards. Our proposed estimator incorporates EOR information resulting in more accurate estimates. Again we define total revenue of 1-C SP t with the ratio estimator $\hat{t}_{\text{ratio}}$ expressed as

$$\hat{t}_{\text{ratio}} = \sum_{i=1}^{5} \hat{t}_{i,\text{ratio}} \tag{3.5}$$

where the ratio estimator

$$\begin{aligned} \hat{t}_{i,\text{ratio}} &= \frac{\sum_{j=1}^{J} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} \frac{x_{ijk}}{n_{ijk}} \sum_{l=1}^{n_{ijk}} y_{ijkl}}{\sum_{j=1}^{J} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}} \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} x_{ijk} && (3.6) \\ &= \frac{\hat{t}_i}{\hat{X}_i} X_i && (3.7) \end{aligned}$$

is used for $i = 1, \ldots, 5$. The proposed estimator is asymptotically unbiased and markedly more efficient than the current production estimator as shown in section 3.3. For detailed description of the estimator, refer to the subsections 3.2.1 - 3.2.5.

3.2.1 Test-Level Notations - (i, j, k) Fixed

For MEP day k in stratum (i, j), we define

$$\begin{aligned} n_{ijk} &= \text{Number of Piece sampled MEP day } k \text{ in strata } (i, j), \\ x_{ijk} &= \text{EOR count for MEP day } k \text{ in strata } (i, j), \\ y_{ijkl} &= \begin{cases} \text{revenue of mailpiece } l, & \text{if } l \text{ in } (i, j, k) \text{ is a 1-C SP}, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The total revenue of 1-C P for test k in stratum (i, j), denoted t_{ijk}, is estimated by

$$\hat{t}_{ijk} = \frac{\sum_{l=1}^{n_{ijk}} y_{ijkl}}{n_{ijk}} x_{ijk}$$

We can express the above as expansion of y_{ijkl},

$$\hat{t}_{ijk} = \sum_{l=1}^{n_{ijk}} u_{ijk} y_{ijkl}$$

where $u_{ijk} = (x_{ijk}/n_{ijk})$ for all $(i, j, k) \in \{1, \ldots, 5\} \times \{1, \ldots, J\} \times \{1, \ldots, n_{ij}\}$. n_{ij} is the number of tests for stratum (i, j) defined in subsection 3.2.2.

3.2.2 Two-Way Strata-Level Notations - (i, j) Fixed

For a given stratum (i, j), we define

$$N_{ij} = \text{Number of MEP days for stratum } (i, j) = (\text{Num. of MEPs}) \times (\text{DelDays}),$$
$$n_{ij} = \text{Number of tests allotted for stratum } (i, j),$$
$$v_{ij} = \frac{N_{ij}}{n_{ij}} = \text{expansion factor for stratum } (i, j).$$

The total revenue of 1-C SP for stratum (i, j), denoted t_{ij}, is estimated by

$$\hat{t}_{ij} = \sum_{k=1}^{n_{ij}} v_{ij}\hat{t}_{ijk}$$

Also, we define the total EOR count for stratum (i, j), symboled X_{ij}, and its estimator, $\hat{X}_{ij}$ as

$$X_{ij} = \sum_{k=1}^{N_{ij}} x_{ijk}$$
$$\hat{X}_{ij} = \sum_{k=1}^{n_{ij}} v_{ij}x_{ijk}$$

3.2.3 One-Way Strata-Level Notations - i Fixed

By summing over SAs, we estimate total revenue of 1-C SP for stratum i, t_i by the ratio estimator of the form

$$\begin{aligned}\hat{t}_{i,\text{ratio}} &= \frac{\sum_{j=1}^{J} \hat{t}_{ij}}{\sum_{j=1}^{J} \hat{X}_{ij}} \sum_{j=1}^{J} X_{ij} \\ &= \hat{R}_i \sum_{j=1}^{J} X_{ij} \qquad \text{for } i = 1, \ldots, 5\end{aligned}$$

The ratio estimator is a “combined ratio estimator” with respect to SA. Since this is done individually for $i = 1, \ldots, 5$, the estimator is a “separate ratio estimator” with respect to BDP categories. By taking the ratio of actual EOR to estimated EOR counts

$$w_i = \frac{\sum_{j=1}^{J} X_{ij}}{\sum_{j=1}^{J} \hat{X}_{ij}} = \frac{X_i}{\hat{X}_i}$$

as an expansion factor we can also express the ratio estimator as an adjustment, through w_i, to the expansion estimator

$$\hat{t}_{i,\text{ratio}} = \sum_{j=1}^{J} w_i \hat{t}_{ij} \qquad \text{for } i = 1, \ldots, 5$$

where $X_i = \sum_{j=1}^{J} X_{ij}$ and similarly for $\hat{X}_i$.

3.2.4 National-Level Notations

National-level total revenue of 1-C SP, t, is therefore obtained by summing individual estimates of BDP categories

$$\hat{t}_{\text{ratio}} = \sum_{i=1}^{5} \hat{t}_{i,\text{ratio}}$$

We can express the estimator as a expansion of original variable y_{ijkl},

$$\hat{t}_{\text{ratio}} = \sum_{i=1}^{5} \sum_{j=1}^{J} \sum_{k=1}^{n_{ij}} \sum_{l=1}^{n_{ijk}} w_i v_{ij} u_{ijk} y_{ijkl}$$

3.2.5 Variance of $\hat{t}_{\text{ratio}}$

An approximate variance of $\hat{t}_{\text{ratio}}$ is obtained by

$$V\left(\hat{t}_{\text{ratio}}\right) = \sum_{i=1}^{5} V\left(\hat{t}_{\text{i,ratio}}\right), \tag{3.8}$$

with

$$V\left(\hat{t}_{\text{i,ratio}}\right) = \left(\frac{X_i}{\hat{X}_i}\right)^2 \left\{\sum_{j=1}^{J} N_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) \frac{S_k^2}{n_{ij}} + \sum_{j=1}^{J} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{N_{ij}} x_{ijk}^2 \left(1 - \frac{n_{ijk}}{x_{ijk}}\right) \frac{S_{kl}^2}{n_{ijk}}\right\} \tag{3.9}$$

and

$$S_k^2 = \frac{1}{N_{ij}-1}\sum_{k=1}^{N_{ij}}\left(t_{ijk} - Rx_{ijk} - (\bar{t}_{ij} - R\bar{x}_{ij})\right)^2$$

$$\bar{t}_{ij} = \frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} t_{ijk}$$

$$\bar{x}_{ij} = \frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} x_{ijk}$$

$$S_{kl}^2 = \frac{1}{x_{ijk}-1}\sum_{l=1}^{x_{ijk}}\left(y_{ijkl} - \bar{y}_{ijk}\right)^2$$

$$\bar{y}_{ijk} = \frac{1}{x_{ijk}}\sum_{l=1}^{x_{ijk}} y_{ijkl}$$

By substituting statistical estimates for the unknown parameters above, we obtain its estimator

$$\hat{V}\left(\hat{t}_{\text{ratio}}\right) = \sum_{i=1}^{5}\hat{V}\left(\hat{t}_{i,\text{ratio}}\right),$$

with

$$\hat{V}\left(\hat{t}_{i,\text{ratio}}\right) = \left(\frac{X_i}{\hat{X}_i}\right)^2\left\{\sum_{j=1}^{J} N_{ij}^2\left(1-\frac{n_{ij}}{N_{ij}}\right)\frac{s_k^2}{n_{ij}} + \sum_{j=1}^{J}\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}} x_{ijk}^2\left(1-\frac{n_{ijk}}{x_{ijk}}\right)\frac{s_{kl}^2}{n_{ijk}}\right\}$$

where

$$s_k^2 = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}\left(t_{ijk} - \hat{R}x_{ijk} - \left(\hat{\bar{t}}_{ij} - \hat{R}\hat{\bar{x}}_{ij}\right)\right)^2$$

$$\hat{\bar{t}}_{ij} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} t_{ijk}$$

$$\hat{\bar{x}}_{ij} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} x_{ijk}$$

$$s_{kl}^2 = \frac{1}{n_{ijk}-1}\sum_{l=1}^{n_{ijk}}\left(y_{ijkl} - \bar{y}_{ijk}\right)^2$$

$$\hat{\bar{y}}_{ijk} = \frac{1}{n_{ijk}}\sum_{l=1}^{n_{ijk}} y_{ijkl}$$

3.3 Comparative Advantage

One obvious advantage of the proposed estimator is attributed to the use of the "correct" day-level expansion factor x_{ijk}/n_{ijk} in (3.6) instead of "estimated" expansion factor $\hat{x}_{jhk}/n_{jhk}$ in (3.2). Quantifying the magnitude of the reduction in bias and variability associated with this change is difficult without conducting an elaborate simulation study, yet the benefit is self-evident.

The second source of improvements comes from the ratio adjustment factor $X_i/\hat{X}_i$ shown in (3.7). Suppose ρ represents the correlation between "EOR count" and "1-C SP revenue," then it can be shown (see [1] on p.157) that the ratio estimator will have a lower variance than the expansion estimator if and only if

$$\rho \geq \frac{1}{2}\frac{(\text{CV of average EOR count})}{(\text{CV of average revenue of 1-C SP})} \tag{3.10}$$

We respectively call the left-hand and the right-hand sides of the inequality "actual correlation" and "critical correlation." Typically, the CV of the EOR count tends to be smaller than CV of 1-C SP volume, because 1-C SP volume depends not only on the EOR count but also on the volume of other products that makes up the EOR count. Then (3.10) suggests that the ratio estimator outperforms the expansion estimator even if the correlation is much lower than 0.5.

The table summarizes the CVs of EOR, 1-C SP volume, and 1-C SP revenue. Based on the CVs shown, we calculate "critical correlations" between EOR and 1-C SP volume, and EOR and 1-C SP revenue for several periods. "Actual correlations" are obtained from ODIS-RPW data for the same periods. Looking at, for example, 1-C SP volume for FY15PQ1 we see a critical correlation of 0.3939. For this product the ratio estimator has smaller variance compared to the current expansion estimator if the correlation between EOR and 1-C SP volume is greater than 0.3939. The table shows that the actual correlation between EOR and 1-C SP volume is 0.6828; implying a significant reduction in variance using the ratio estimator. In fact, all periods and variables, we consistently observe the advantage of the ratio estimator in lowering variances.

Using the same data, we predict that CVs for 1-C SP volume is reduced by 15-20% and those for 1-C SP Revenue by 5-10%. While not shown, the computations are from comparisons of first-stage variances between the current in (3.4) and the proposed ratio estimator in (3.9).

Finally, although not quantified, we expect some reduction in variance through stratification based on BDP.

Period	Variable	CV	Cri. Corr.		Act. Corr.
FY15 PQ1	EOR	0.0075			
	Piece	0.0095	0.3939	$<$	0.6823
	Revenue	0.0267	0.1406	$<$	0.1864
FY15 PQ2	EOR	0.0080			
	Piece	0.0104	0.3875	$<$	0.7059
	Revenue	0.0132	0.3047	$<$	0.4798
FY15 PQ3	EOR	0.0080			
	Piece	0.0105	0.3787	$<$	0.5202
	Revenue	0.0124	0.3221	$<$	0.4991

Table: Comparisons of Critical and Actual Correlation by Period. The ratio estimator has a lower variance if Cri. Corr. $<$ Act.Corr. Regardless of variables estimated, pieces or revenue, Act. Corr is large enough to conclude that the ratio estimator produces estimates with lower variances.

References

[1] Cochran, William G. (1977), *Sampling Techniques*. 3rd ed., John Wiley & Sons (Hoboken, NJ).